DOCUMENT RESUME

ED 372 085 TM 021 726

AUTHOR Kluever, Raymond C.; Green, Kathy E.
TITLE An Examination of Ethnic and Gender Differences in the Raven Coloured Progressive Matrices Test.
PUB DATE Apr 94
NOTE 21p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Anglo Americans; *Cultural Differences; Elementary Education; Elementary School Students; *Ethnic Groups; *Hispanic Americans; *Intelligence Tests; Rural Schools; *Sex Differences; Test Bias; Test Results; Test Use
IDENTIFIERS *Ravens Coloured Progressive Matrices

ABSTRACT

Response patterns to the Raven Coloured Progressive Matrices (CPM) were analyzed for a sample of 203 Hispanic and 254 Anglo first- through fifth-grade children from a rural school district in southern Colorado. Gender distributions were nearly equal. Gender and ethnic differences were examined within the context of determining whether the CPM functioned in a similar manner across groups. The greatest differences in test structure were found between boys and girls, not between ethnic groups. Some performance differences were found between Anglo and Hispanic children, particularly with fourth- and fifth-grade girls. Results suggest that use of the CPM subtests may provide insight into the mechanisms underlying these gender and ethnic differences. One figure and eight tables are included. (Contains 10 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 372 085

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
RAYMOND C. KLUEVER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

An Examination of Ethnic and Gender Differences in the Raven Coloured Progressive Matrices Test

Raymond C. Kluever, Ph.D.
Associate Professor, Emeritus
University of Denver

and

Kathy E. Green, Ph.D.
Associate Professor
University of Denver

Paper presented at the annual meeting of the American Educational Research Association, New Orleans, April 1994

TM021726

BEST COPY AVAILABLE

Abstract

Response patterns to the Raven Coloured Progressive Matrices (CPM) were analyzed for a sample of 203 Hispanic and 254 Anglo 1st - 5th grade children from a rural school district in southern Colorado. Gender and ethnic differences were examined within the context of determining whether the CPM functioned in a similar manner across groups. The greatest differences in test structure were found between boys and girls; not between ethnic groups. Some performance differences were found between Anglo and Hispanic children, particularly 4th and 5th grade girls. Results suggest that use of the CPM subtests may provide insight into the mechanisms underlying these gender and ethnic differences.

Acknowledgement

We are grateful for the assistance provided by Mr. Gary Shawcraft, Assistant Superintendent in the North Conejos School District in LaJara, Colorado. His interest in and contribution to this project were invaluable.

Concern about ethnic, cultural, and gender differences in tests and test items has been an issue since the 1950s. A substantial literature covers this topic and it is growing. Sattler (1988) provided a discussion of the major criticisms of the use of tests with ethnic minority children. Some of the criticisms are:

> -Intelligence tests have a cultural bias.
> -National norms are inappropriate for minorities.
> -Minorities are handicapped in test taking skills.
> -The fact that most examiners are white has the effect of depressing the scores of ethnic minority children.
> -Test results lead to inadequate and inferior education. (p. 566).

Ethnic, cultural, and gender differences have led to investigations of potential test bias. Although bias and misuse of tests has occurred and has been delineated by critics, supporters of test use have provided ample bases for legitimate test use. Shepherd (1981) noted that "The defenders of tests acknowledge specific instances of cultural bias in some tests but do not believe that the use of tests is necessarily or uniformly unfair." (p. 79).

A number of statistical methods to probe potential internal sources of bias have been proposed. Reynolds (1982) noted the primary ones as "equivalent internal consistency reliability estimates, a high level of factorial similarity" and "systematic error in the estimation of the criterion score as a function of group membership" (p. 224). Mercer (1984) listed 4 characteristics of a racially and culturally nondiscriminatory test as it related to her work.

> 1. equal internal integrity.
> 2. equal external relevance for the groups on whom the procedures are to be used.
> 3. equivalent item patterning.
> 4. equal intercorrelational and/or factor patterns for all the groups on which it is to be used. (p. 296)

She presented an extensive elaboration of each of these four characteristics and methodology for using them to minimize the occurrence of bias in tests. A review of more recent statistical methodologies and practical concerns related to test bias may be found in Holland and Wainer (1993).

Intelligence is currently conceptualized as a composite of many abilities which are organized into subtests on commonly used tests of intelligence. When these tests are used with different ethnic groups, differences in mean level of performance can be observed. Some group's mean performance is above the norm and other group's mean performance is below the norm. The use of "cut-off" scores as the major criterion for educational decision

making had led to a disportionately larger number of children from some ethnic groups being placed in special educational programs while the number of children from other groups was below expectation. The curriculum in these special programs and the stigma attached to such labeling and placement have been a major concern (Mercer, 1979).

A review of the Raven Progressive Matrices Tests suggests that this test may represent a measure with less variability among ethnic groups than is found in some of the commonly used tests. The authors (Raven, J., Raven, J.C., & Court, H.H., 1991) described the theoretical basis for the test as follows.

> Raven's Progressive Matrices and Vocabulary tests were developed to assess, as simply and unambiguously as possible, the two components of g identified by Spearman as eductive ability and reproductive ability. (p. G1).

and that, "The Matrices measure the ability to educe relationships" (p. G3). Eductive ability, "is the process of educing, or squeezing, new insights and information out of that which is perceived or already known" (p. G2).

Three versions of the Progressive Matrices tests exist: the Advanced Progressive Matrices Test (APM), the Standard Progressive Matrices Test (SPM), and the Coloured Progressive Matrices Test (CPM). The items consist of designs with the examinee required to select one of 6 or 8 choices that is the best fit for a missing area of the major design. The CPM test has colored items and is appropriate for elementary school children. The others have black and white designs and are more appropriate for older subjects.

The purpose of this study was to examine ethnic and gender differences in CPM item functioning, total score, and in internal characteristics of the test. The focus of this study was on overall and subtest functioning rather than item performance. The relationship of CPM scores to achievement was also examined since ability is frequently associated with high and low achievement scores.

Method

The setting for this study was a rural southern Colorado school district where agriculture is the major industry. All 471 children in grades one through five in the two elementary schools were tested by the School District Assistant Superintendent in the fall of 1992. The test was presented in class as a game or puzzle rather than as an achievement test. The sample was composed of 254 Anglo children (57%) and 203 Hispanic children (43%). The gender distribution was nearly equal (50.3% female, 49.7% male). Enrollment by grade was 80 first graders, 105

second graders, 89 third graders, 100 fourth graders, and 97 fifth graders.

The Iowa Test of Basic Skills had been administered the previous spring (1992) as part of the school district testing program. A few students who were new to the school district did not have ITBS scores but their CPM scores were available. The ethnicity and gender of 14 children was not identifiable from the records and these children were omitted from analyses.

Results

Statistically significant differences were found on the CPM total raw score and the ITBS composite percentile for Anglo and Hispanic children, ($t=-2.90$, $p<.004$). Anglo children had a mean score of 27.1 CPM items correct and Hispanic children had 25.4 items correct; a difference of 1.7 items (Table 1). The effect size (.28) was small. The CPM mean raw scores increased from a mean of 20.4 items correct for 6 year olds to 29.8 items correct for 11 year olds. The scores for 10 and 11 year olds were not significantly different ($t=.62$, $p>.05$) suggesting that beyond 10 years of age, the difficulty level of the test is such that it may no longer discriminate among children of increasing age. A summary of the mean raw scores for each age and for each ethnic group suggests that there is a pattern of consistent increase in score with increase in age (Table 1). Seven year old Anglos were the exception. The mean score for the Hispanic children appears to be about one year delayed over that of comparable age Anglo children. There was no change in the difference between Anglo and Hispanic mean scores throughout the age span of 6 through 11 years.

The internal consistency reliability of the CPM was .92. The reliability estimates for each ethnic group were the same, .88. For males, the reliability was .90 and for females, .87.

The difference between the means of the Anglo and Hispanic ITBS percentile score was 15 with a medium effect size (.56). The mean percentile score was 63.4 for Anglo children and 48.4 for Hispanic children (Table 1).

There was no statistically significant gender difference between CPM total scores ($t=-.98$, $p>.05$) but there were significant differences on the ITBS composite percentiles ($t=-2.41$, $p<.05$). The mean percentile for males was 53.7 and for females 61.1.

The factor structure on the CPM for Anglo and Hispanic children was was very similar to that of the total group. (Tables 2 & 3). A principal components analysis with varimax rotation and loadings of .4 or higher were used in defining each factor (Table 2). Three factors were identified for the total group and the same ones for each ethnic group. They were interpreted to be a

matching task, a completion/closure task, and an analogies task. Four items, A8, A10, AB10, and AB11 did not load on any factor and 1 item double loaded. Forty-nine percent of the variance was accounted for in this analysis. In the analysis by each ethnic group, the 3 factors accounted for 34% of the variance for both Anglo and Hispanic children.

The CPM factors extracted for males and females were different (Table 4). Factor 1 for males was a figure completion/closure task but for females, it was a figure matching task. The second factor was an analogies factor for males but a figure completion and closure factor for females. And, the third factor was interpreted to be a matching factor for males and an analogies factor for females. These three factors accounted for 47% of the variance for females and 35% for males. These findings suggest differences in CPM factor structure for males and females.

Then, using the three factors identified by the composite analysis using all cases as 3 subtests of the CPM and the number of items correct on each factor as the "subtest score," significant ethnic differences were found only on factor 2, the completion/ closure factor. Performances on the matching factor and the analogies factor were essentially the same (Table 5). The internal consistency reliabilities of the 3 subtests for Anglos were .43, .84, and .73 for factors (subtests) 1, 2, and 3. For Hispanic children the reliabilites were .57, .83, and .72.

The mean subtest pattern for males and females was the opposite of the pattern for ethnicity (Table 5). That is, there were significant differences on subtest 1 and subtest 3 but not on subtest 2. In all cases, females had higher scores; significantly higher for factors 1 and 3. However, the pattern of correlations between subtests was similar for the two groups. For females, the correlations varied from .23 between factors 1 and 3 to .64 between factors 2 and 3 ($p < .01$). And, for males, the correlation varied from .32 between factors 1 and 3 to .60 between factors 2 and 3 ($p < .01$). It was reported above that the mean CPM score did not differ significantly between males and females but the internal configuration of the test did show gender variation.

A stepwise multiple regression equation was computed for each ethnic group. The ITBS composite total percentile served as the dependent variable and CPM total and item scores, grade, and gender were the dependent variables. The R values for each group were identical (.57) but there were small differences in the combination of specific CPM items that entered the regression equation as predictors (Table 6). The total CPM score, gender, and the child's grade were significant contributors to determining the value of R. Their order of entry into the equation was the same for both groups.
The prediction of ITBS total composite score for males and

females resulted in an R value of .62 for males (R sq.=.39) and .56 for females (R sq.=.31). Only the grade in which the student was enrolled was a significant predictor for both groups (Table 7). For males, the student's grade, the CPM total score and item 35 were the best predictors while for females, grade and a series of 5 CPM items were the best predictors. For females, the CPM total score entered the equation but was removed a few steps later. The predictors and the order of their entrance into the regression equation were similar for males and for the total group (males and females combined) but differed for females.

A discriminant analysis was used to predict group membership (Anglo or Hispanic) from the CPM items. For Hispanic children, the pattern indicated that these children will be predicted to belong to the Hispanic group about half time and that they have patterns resembling Anglo children half of the time. The pattern for Anglo children is somewhat more specific in that Anglo children are predicted to be Anglos about three fourths of the time. On the basis of CPM item responses, it is very difficult to identify children as having a clearly identifiable Anglo or Hispanic pattern of performance (Table 8).

The discriminant analysis predicting group membership by gender was similar to that for the ethnic grouping (Table 8). The CPM pattern of scores for females was predictive of their gender group 2/3 of the time and males 1/3 of the time. These findings suggest that prediction of gender from CPM scores would result in a large margin of error.

The analyses reported above are based on the use of traditonal test theory to create test and subtest scores. An analysis based on item response theory was also conducted. CPM item responses were analyzed for fit to a 1-dimensional model using BIGSTEPS (Wright & Linacre, 1993). Data were analyzed both jointly and separately for Anglo and Hispanic boys and girls. Misfitting items and items that could not be calibrated because they were too easy for a subgroup were omitted, leaving 23 useful items. Anomalous responses (e.g. a child missing many easy items and correctly answering a very difficult item) were treated as missing data; 2 cases were deleted (1 Anglo and 1 Hispanic boy) in which responses appeared to be random. Items remaining were then calibrated in units called logits (log-odds units) for the entire sample and separately for each subgroup. Logits typically range from -3 to +3 with items set to a mean value of 0.0.

Analyses were conducted separately to see if the variable retained it's definition for each subgroup and to see if items functioned differently for different subgroups. The correlations between item logit values for Anglo boys and girls, Hispanic boys and girls, and Anglo and Hispanic boys were all r = .98; for Anglo girls and Hispanic girls the correlation was r = .97. These very high correlations suggest that the items function in a

strikely similar manner for all 4 subgroups. One item was significantly easier for Hispanic boys than for Anglo boys (B10); 2 items were easier for Anglo girls than Hispanic girls (A7, A9) while 1 item was easier for Hispanic than Anglo girls (A12). These differences were minor.

Greater differences were found between boys' and girls' performance for both Anglos and Hispanics. For both groups, some items from the first set (Set A) seem to favor boys while some items for the last set (Set B) favor girls. This supports the finding reported earlier that girls and boys differ in performance level on subparts of the test, boys have an advantage in pattern matching and girls in analogies.

While statistically significant differences were found overall between Anglo and Hispanic students, when analyzed by subgroup it was noted that this difference was due to the performance of Anglo and Hispanic girls, not boys. The performance of Anglo and Hispanic boys did not differ nor did boys' mean scores differ from girls' mean scores. Hispanic girls had the lowest scores and Anglo girls the highest ($t=2.76$, $p<.01$). To further investigate this difference, scores for each subgroup were broken down by grade. Figure 1 presents the mean logit values for the 4 gender/ethnic subgroups by grade; 95% confidence bands are also drawn in the figure. Scores for all groups increase substantially between grade 2 and grade 3. But, for Anglo girls and boys and Hispanic boys, the gain after grade 2 was larger than for Hispanic girls. While this result needs to be replicated for other samples, it suggests that there are only minor differences in performance on the CPM between Anglo and Hispanic boys and girls through grade 5 and that ethnic differences found may be gender specific.

Discussion

The purpose of this study was to analyze data that would provide some preliminary evidence concerning the presence or absence of ethnic and/or gender differences in the functioning of the CPM test. The evidence was based on an analysis of the internal characteristics of the CPM; that is, the reliability, factorial validity, and overall test structure.

Statistically significant differences were found between scores of the Anglo and Hispanic children on the CPM total score. The effect size was small (.28) and the difference was driven by a difference in performance level between Anglo and Hispanic girls. Although the total score difference was statistically significant, it is not an indication of a test that is biased against Hispanic children. In fact, the test functions in a similar manner for Anglo and Hispanic children. Raven and Court (1989), citing Felmlee's study of a similar sample of Anglo and Hispanic children from the same location, stated that "the

results were so similar that it has not been necessary to publish separate norm tables" (p rs4.14).

No significant gender differences were found between male and female total CPM scores but performance for boys and girls differed.

The reliability of the CPM (Cronbach's Alpha) was the same for the Anglo and Hispanic groups (.88). Small differences were noted for the reliability of males and females (.90 and .87). These values suggest that the test is equally reliable for both ethnic and gender groups. Group differences in test reliability (internal consistency) have been regarded as evidence of test bias but Bond (1982) regarded reliability as a weak indicater of bias since a test can be invalid but possess high reliability.

The factorial validity of the test indicated clearly that the test is measuring the same construct for each ethnic group. In fact, the pattern of items that loaded on each factor was nearly the same as that for the combined Anglo and Hispanic groups. But, the analysis by gender resulted in different male/female factor patterns (Table 4). Factor 1 for females included items representing figure matching tasks whereas the items for factor 1 for males involved figure completion/closure items. Factor 2 for females involved a figure completion/closure task but for males, it was made up of items involving analogies. Figure matching was the primary task required for males on factor 3 but for females it was an analogies task. The pattern of loadings on factors for females shows more variation in the types of items that make up each factor than was found for males. Males have a clearly interpretable pattern of item loadings for each factor. The percent of variance of each factor also varies by gender. That is, the figure matching task involves 31.4 percent of the variance for females but only 6.4 percent for males. The remaining two factors show similar differences by gender. The nature of the gender differences are not clear from this data set and should be explored further. The evidence suggests that the structure of the CPM is different for males and females.

The total CPM score is nearly as predictive of the composite ITBS score for Hispanic children (R sq. = .09) as for Anglo children (R sq. = .11). When three other variables are included in the equation, the R square value becomes the same for each of them. The great similarity of the variables in the prediction equation and the very similar R square values suggests that the CPM is not a biased predictor of ITBS scores for these two ethnic groups.

But the prediction of achievement for males and females based on the CPM values is very different. In a multiple regression equation which included the total CPM score along with all of the CPM items and the subject's grade in school, the total CPM score became the best predictor for males. Grade and one CPM item

entered the equation later. However, for females a series of CPM items and the student's grade were the best predictors of ITBS achievement. The total CPM score entered the equation but was removed at a later step. These findings indicate the the CPM shows equivalent but differentially structured prediction of ITBS achievement scores for males and females.

A discriminant analysis were used to indicate whether each ethnic group or whether males and females could be identified by their pattern of responses on the CPM. That is, does each group have a distinct CPM pattern which can be clearly differentiated from the pattern for the other group? The results of the discriminant analysis suggest that it is very difficult to make this distinction. These results indicate that the groups demonstrate similar performance patterns on the CPM and therefore that the CPM measures the same variable for these groups.

Finally, the analysis via item response theory found no evidence of differential test functioning by ethnic group. Greater evidence of differential functioning was found by gender.

This analysis of possible ethnic and gender differences in the CPM indicates that it is probably not an ethnically biased test when used with Anglo and Hispanic children. The group mean scores are different, but the similarity between the reliability values, the item characteristics, the factorial validity, the regression analysis for predicting academic achievement, and the regular increments in CPM mean scores with age are evidence of the internal characteristics of a nondiscriminatory test. But, analysis relating to gender indicated some evidence for differential test functioning between boys and girls for both Anglos and Hispanics. No evidence of gender differences in CPM total score differences were found and the test reliabilities were nearly the same for each sex. The discriminant analysis is a further indication of similarity of item patterns for males and females. However, the regression equations for predicting achievement were different, the validity was different, and some items functioned differently. These are significant indicators of the internal characteristics of test bias and would suggest that the CPM possesses some gender specific structure. Further analysis of the possible gender bias in the CPM is recommended. A different sample of subjects, analysis of bias for each ethnic group separately, and analysis by grade level may clarify when and where differences appear. Results here suggest that differences appear at grade 4.

This preliminary analysis provides an initial indication of ethnic and gender differences in the CPM. Further analyses are recommended and would include evidence of test-retest reliability and a further study of predictive validity for specific purposes. However, the evidence accumulated would suggest that the test is to be recommended when possible ethnic bias is a concern but

gender differences must be considered in the use of the test.

The finding of significant differences in subtest performance leads one the question whether there are different gender related progressions in the development of some specific skills (e.g. analogies). Future investigations might profit from inclusion of specific measures of matching and analogies performance.

Figure 1

Anglo and Hispanic Girls and Boys Mean Logit Value by Grade

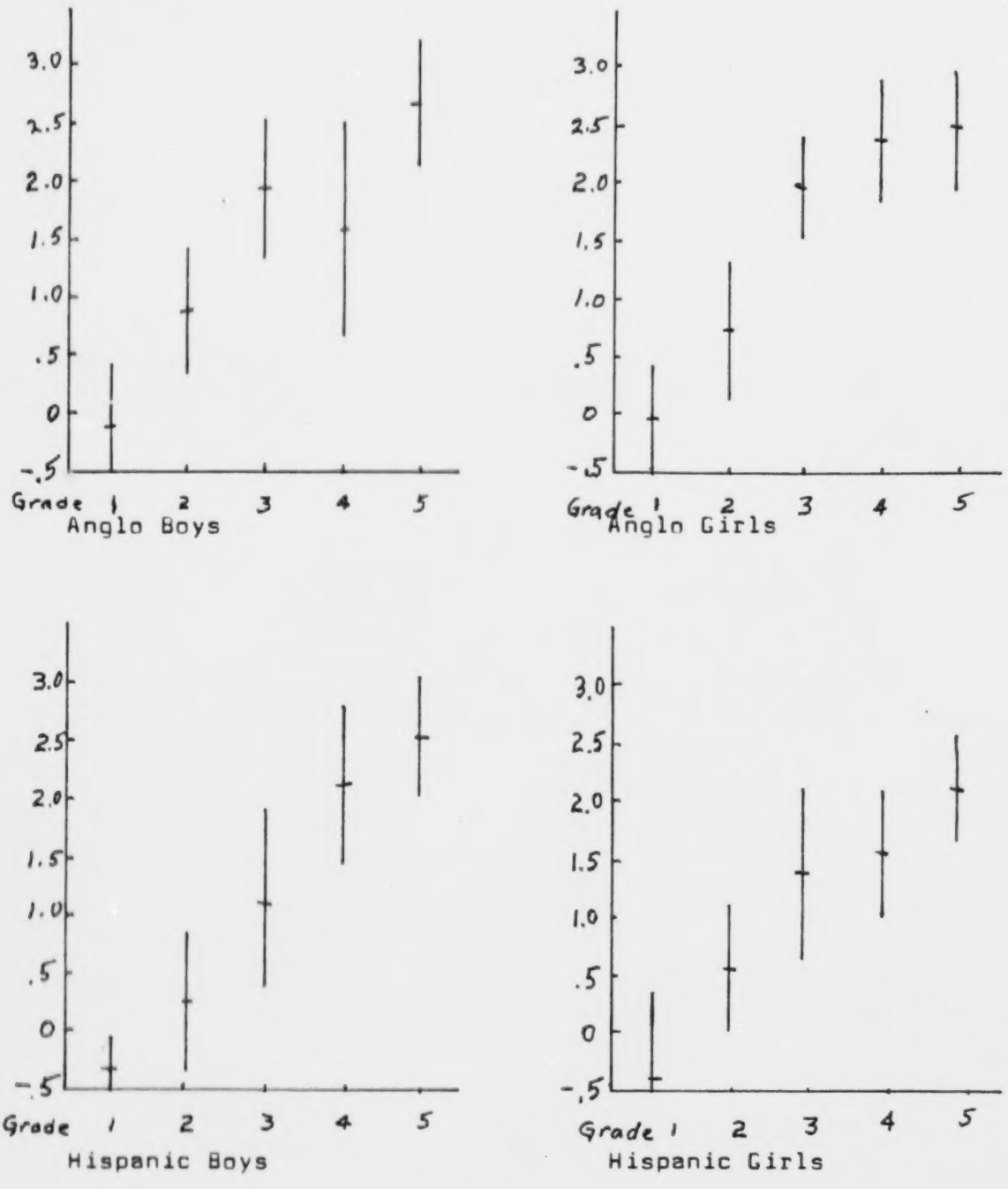

Table 1

Mean CPM and ITBS Scores of Anglo and Hispanic Children by Age and Gender

	Anglo				Hispanic			
Variable	N	Mean	SD	Range	N	Mean	SD	Range
6 yr CPM	22	22.4	4.5	14-31	16	17.7	5.0	12-30
7 yr CPM	46	21.7	5.1	13-34	44	22.3	6.1	13-35
8 yr CPM	54	26.6	5.6	10-35	43	24.1	5.6	11-35
9 yr CPM	50	28.6	4.7	13-35	38	27.8	4.3	20-36
10 yr CPM	49	31.3	4.2	16-36	39	28.8	4.7	13-36
11 yr CPM	32	30.1	3.9	22-36	21	29.3	3.6	20-34
Tot. CPM	254	27.1	5.9	10-36	203	25.4	6.1	11-36
CPM Fem.	133	27.4	5.7	13-36	97	25.6	5.5	15-36
CPM Male	121	26.7	6.2	10-36	106	25.3	6.7	11-35
Tot. ITBS	221	63.4	24.8	1-99	184	48.4	28.4	1-99
ITBS Fem.	115	66.5	21.6	11-99	85	50.7	26.8	1-99
ITBS Male	106	60.1	27.6	1-99	99	46.5	29.8	1-99

Table 2

Factor Loadings of the CPM Items for all Groups (N-471)

Item	Factor 1 (Matching)	Factor 2 (Completion/ Closure)	Factor 3 (Analogies)
A1	.94		
A2	.94		
A3	.88		
A4	.85		
AB1	.85		
A5	.79		
AB2	.77		
B1	.73		
A6	.73		
AB3	.68		
B2	.66		
AB6		.65	
AB4		.63	
A7		.62	
B5		.61	
AB5		.61	
AB7		.60	
AB8		.59	
AB21		.58	
B4		.57	
B3	.46	.52	
AB10		.45	
A9		.44	
AB11		.43	
B6		.41	
A?			.77
B11			.76
B8			.71
B10			.69
AB12			.61
A12			.55
Eigenvalue	11.85	4.22	1.68
% Variance	32.9	11.7	4.7
% Cum. Var.	32.9	44.6	49.3

Table 3

Factor Structure for Anglo and Hispanic Groups

	Matching Factor		Completion/ Closure Factor		Analogies Factor	
	Ang.	Hisp.	Ang.	Hisp.	Ang.	Hisp.
	A3	A3	A7	A7	A11	A11
	A4	A4	A9		A12	
	A5	A5	AB1	AB1		AB6
	A6	A6	AB2	AB2		AB8
		B1	AB3	AB3	AB9	AB9
			AB4	AB4	AB12	AB12
			AB5		B5	
			AB6	AB6	B6	
			AB7	AB7	B7	B7
			AB8		B8	B8
			B2	B2	B9	B9
			B3		B10	B10
			B4	B4	B11	B11
			B5	B5	B12	B12
Eigenvalues	7.26	7.25	2.29	2.24	2.11	2.20
% Variance	21.4	21.3	6.7	6.6	6.2	6.5
% Cum. Var.	21.4	21.3	28.1	27.9	34.3	34.4

Table 4

Factors and Factor Loadings of CPM Items by Gender

	Females			Males		
Item	Fact 1	Fact 2	Fact 3	Fact 1	Fact 2	Fact 3
A1						
A2						
A3	.88					.54
A4	.85					.68
A5	.81					.82
A6	.74					.68
A7		.62		.58		
A8						
A9		.43		.45		
A10						
A11					.44	
A12					.48	
AB1	.84			.54		
AB2	.78			.50		
AB3	.69			.59		
AB4		.63		.67		
AB5		.62		.49		
AB6		.65		.60		
AB7		.59		.63		
AB8		.59		.46		
AB9		.58		.42	.49	
AB10		.45				
AB11		.43		.45		
AB12			.61		.62	
B1	.74					
B2	.68			.55		
B3	.46	.52		.52		
B4		.56		.64		
B5		.61		.59		
B6		.41			.41	
B7						
B8			.71		.67	
B9			.77		.73	
B10			.69		.71	
B11			.77		.77	
B12			.55		.53	
Eigenvalues	10.68	3.65	1.65	8.18	2.50	2.16
% Variance	31.4	10.7	4.8	24.0	7.4	6.4

Table 5

Mean Gender and Ethnic Scores on the Three Factors (Subtests)

Factor	Gender/Ethnicity	Mean	SD	t	p
1	Female	11.83	.46	3.34	<.01
	Male	11.59	.94		
2	Female	10.35	3.13	0.36	>.05
	Male	10.23	3.68		
3	Female	2.69	2.06	2.03	<.05
	Male	2.29	2.11		
1	Anglo	11.74	.68	0.91	>.05
	Hispanic	11.67	.82		
2	Anglo	10.68	3.33	2.72	<.01
	Hispanic	9.81	3.46		
3	Anglo	2.66	2.12	1.93	>.05
	Hispanic	2.28	2.03		

Table 6

Stepwise Multiple Regression Analysis Using ITBS Composite Score as the Dependent Variable for Anglo and Hispanic Groups Separately

	Anglo			Hispanic		
Step	Variable Entered	R	R sq. Adj.	Variable Entered	R	R sq. Adj.
1	CPM Total Score	.34	.11	CPM Total Score	.29	.09
2	Grade	.53	.28	Grade	.55	.30
3	Gender	.55	.30	Gender	.56	.31
4	Item A4	.57	.32	Item B11	.57	.32

Table 7

Stepwise Multiple Regression Analysis by Gender

	Females			Males		
Step	Variable Entered	R	R sq.	Variable	R	R sq. Entered
1	Item B11	.31	.09	Total Score	.31	.10
2	Grade	.43	.18	Grade	.61	.38
3	Total Score	.50	.25	Item B9	.62	.39
4	Item A10	.52	.27			
5	Item AB8	.54	.29			
6	Total Score (Removed)	.53	.28			
7	Item B10	.55	.30			
8	Item A5	.56	.31			

Table 8

Discriminant Analysis Group Classification by Ethnicity and Gender

Actual Group	Predicted Group	
	Anglo	Hispanic
Anglo	76%	24%
Hispanic	52%	48%

Percent of cases correctly classified by ethnicity = 64%

	Female	Male
Female	65%	35%
Male	41%	59%

Percent of cases correctly classified by gender = 62%

References

Bond, L. (1981). Bias in mental tests. In B.F. Green (Ed.). Issues in testing: coaching, disclosure, and ethnic bias (p. 55-78). San Francisco: Jossey-Bass.

Holland, P.W., & Wainer, H. (Eds.) (1993). Differential item functioning. Hillsdale, NJ: Lawrence Erlbaum Associates.

Mercer, J.R. (1979). System of multicultural pluralistic assessment: technical manual. New York: The Psychological Corp.

Mercer, J.R. (1984). What is a racially and culturally nondiscriminatory test? A sociological and pluralistic perspective. In C.R. Reynolds & R.T. Brown (Eds.). Perspectives on bias in mental testing (p. 293-356). New York: Plenum Press.

Raven, J.C., & Court, J.H. (1989). Manual for the Raven's Progressive Matrices and Vocabulary Scales; Research Supplement No. 4. London: H.K. Lewis & Co. Ltd.

Raven, J.C., Court, J.H., & Raven J.H. (1991). Manual for Raven's Progressive Matrices and Vocabulary Scales, Section 1, General Overview. Oxford: Oxford Psychologists Press.

Reynolds, C.R. (1982). Methods for detecting construct and predictive bias. In R.A. Berk (Ed.). Handbook of methods for detecting item bias (p. 199-227). Baltimore: Johns Hopkins University Press.

Sattler, J.M. (1988). Assessment of children (Third Edition). San Diego: J.M. Sattler, Publisher.

Shepard, L.A. (1981). Identifying bias in test items. In B.F. Green (Ed.). Issues in testing: coaching, disclosure, and ethnic bias (p. 79-104). San Francisco: Jossey-Bass.

Wright, B.D. & Linacre, J.M. (1993). A user's guide to BIGSTEPS. Chicago: MESA Press.